LEVEL II (12)

ACCESSION for
NTIS White Section
DDC Buff Section
A

(6) GOODNESS OF FIT TESTS WITH SPECIAL REFERENCE TO TESTS FOR EXPONENTIALITY.

BY

(10) Michael A. Stephens

(9) TECHNICAL REPORT NO. 262

(11) SEPT[redacted]978 (12) 32p.

Prepared under Contract
N00014-76-C-0475 (NR-042-267)
For the Office of Naval Research

(14) TR-262, TR-22

Herbert Solomon, Project Director

AD A060824

DDC FILE COPY

(15) N00014-76-C-0475, DAAG29-77-G-0081

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

Partially supported under U.S. Army Research Office Grant DAAG-29-77-G-0031
and issued as Technical Report No. 22.

78 11 01 053 JCB

332 580

GOODNESS OF FIT TESTS

WITH SPECIAL REFERENCE TO TESTS FOR EXPONENTIALITY

1. In recent years there have been big advances in goodness-of-fit testing; famous tests like X^2 and those based on the empirical distribution function or on spacings have been improved or extended, particularly to deal with nuisance parameters; new tests have been proposed for special situations like testing for normality and exponentiality; modern computer techniques have enabled new examinations of older statistics like b_1 and b_2 measuring skewness and kurtosis.

We wish to give a guide to some of these advances, following some principles which we hope are based on practical considerations.

2. The problem is as follows: a random sample of n values of x is given, and we wish to test the null hypothesis, that the distribution function of x, $G(x)$, is a given distribution

$$H_0: G(x) = F(x;\underset{\sim}{\theta})$$

where $\underset{\sim}{\theta}$ is a vector of parameters which may be partly or wholly unknown.

3. Historically, the classical test is Pearson's X^2, usually called chi-square; although of long history, much work is still being done on this statistic. It is a natural statistic for discrete distributions, but when $F(x;\underset{\sim}{\theta})$ is continuous, one loses information by grouping; however, the test is easily adapted for use when $\underset{\sim}{\theta}$ is unknown, or part of $\underset{\sim}{\theta}$ is unknown. The adaptation implies a method of estimating parameters which

is often not followed in practice, and also uses asymptotic results for finite samples, and errors are introduced in this way; nevertheless, the statistic is well established in the user's lexicon.

4. There are two other assets for χ^2 which can be turned into principles for practical use.

Principle 1. The user likes to keep close to the original data; he does not want, if it is avoidable, to make an elaborate transformation of his x-values to, say, new z-values, which are then used in calculating a test statistic, but which as a data set mean little to him.

Principle 2. When a test statistic is significant, the user will want to interpret this in terms of some irregularity in the original data, which is suggested by the form of the test statistic.

In the χ^2 test, one sees the original data on a line, and only grouping is involved; thus Principle 1 is observed. With reference to Principle 2, a high value of the statistic can usually be seen to come from one or two cells and the irregularity in the data is at once pinpointed.

5. Over the years, other methods of testing have been introduced. An important problem is always how to handle nuisance parameters. We shall discuss only tests for continuous distributions, for which chi-square is not so naturally suited, and will begin with two general types of test: EDF statistics and regression tests.

6. EDF statistics are the oldest historically; the first was the Kolmogorov (-Smirnov) statistic. EDF refers to the empirical distribution function $F_n(x)$:

$$F_n(x) = \frac{\text{number of x-values} \leq x}{n} ,$$

and EDF statistics are based on a measure of the discrepancy between $F_n(x)$ and $F(x;\theta)$. These are both graphed; and, loosely speaking, D, the Kolmogorov-Smirnov statistic, is the largest vertical gap between the curves. More precisely, let $F(x)$ refer to $F(x;\theta)$ with θ known, or with estimates of θ inserted where necessary. Then

$$D^+ = \sup_x \{F_n(x) - F(x)\}, \; D^- = \sup_x \{F(x) - F_n(x)\}$$

and $D = (D^+, D^-)$.

A related statistic is (Kuiper's) $V = D^+ + D^-$.

Other statistics measure the discrepancy in other ways; three of the most important are the Cramér-von Mises W^2, Watson U^2, and Anderson-Darling A^2:

$$W^2 = n \int_{-\infty}^{\infty} \{F_n(x) - F(x)\}^2 \, dF(x) ,$$

$$U^2 = n \int_{-\infty}^{\infty} \{F_n(x) - \int_{-\infty}^{\infty} (F_n(y) - F(y)) \, dF(y)\}^2 \, dF(x)$$

$$A^2 = n \int_{-\infty}^{\infty} w(x) \, \{F_n(x) - F(x)\}^2 \, dF(x) ,$$

with

$$w(x) = \frac{1}{F(x)\{1 - F(x)\}} .$$

In order to fix our ideas, and because we wish to concentrate later on this test, let us suppose that the hypothesis to test is that n values of x constitute a random sample from the exponential distribution:

$$F(x;\theta) = 1 - \exp(-\beta(x-\alpha)) , \quad x > \alpha ;$$

$\underset{\sim}{\theta} = (\alpha,\beta)$ and one or both may be unknown. When $\underset{\sim}{\theta}$ is known, asymptotic and much finite-n null distribution theory is known for the above statistics, and they have slowly come into use. Undoubtedly, wider use was held up because in the important situations where $\underset{\sim}{\theta}$ was not known, the distribution theory was lacking. In very recent years this has been somewhat filled in, and EDF statistics can now be used for tests for the following important distributions:

(a) the normal, with μ and/or σ^2 unknown,

(b) the exponential distribution, β unknown, α known.

It is of course easy to test for the exponential distribution when α is also not known; the smallest number of the sample is subtracted in turn from all the others and the resulting set is tested for exponentiality with β unknown and α equal to zero. The above tests are in the literature, and are described in Stephens (1974b). Also, significance points have recently been provided for tests for:

(c) the Gamma distribution, scale parameter unknown (Pettitt and Stephens, 1976),

(d) the extreme value distribution, scale and/or location parameters unknown (Stephens, 1976b).

In summary of the theoretical side, asymptotic distributions of EDF statistics do not depend on $\underset{\sim}{\theta}$ if this involves location and scale parameters. Asymptotic points have been calculated for W^2, U^2 and A^2 (Stephens, 1976a; Durbin, Knott and Taylor, 1975), and recently progress has been made for finite n, for the test for exponentiality, by Durbin (1975) and Margolin and Maurer (1976). But in general, points for all statistics, for finite n, have been found by Monte Carlo methods. These must be extrapolated to give asymptotic points for $\sqrt{n}\,D$ and $\sqrt{n}\,V$. For all the statistics Stephens (1974b; also given in Table 54 of Pearson and Hartley, 1972) has provided correction factors which enable the statistics to be used with only the asymptotic points. Many other references to work in this area are in Stephens (1974b, 1976a); for an overall summary of the EDF and its properties see Durbin (1973).

Advantages of EDF statistics are:

(a) they follow Principles 1 and 2;

(b) they are easily computed (one formula for all n);

(c) over a wide range of alternatives they are more powerful than χ^2 (probably because of grouping in χ^2);

(d) they provide <u>consistent</u> tests; if x in fact comes from $G(x)$ which is not $F(x;\theta)$, then as $n \to \infty$, $F_n(x) \to G(x)$, and the test statistics will be <u>declared significant</u>.

Points (c) and (d) can be included as two new principles:

<u>Principle 3</u>. A test should be <u>consistent</u> and <u>unbiased</u>.

<u>Principle 4</u>. A test should be powerful over a wide range of alternatives.

Clearly we cannot find a test which is, in general, most powerful (or even powerful) over all alternatives to $F(x;\underset{\sim}{\theta})$; and since the alternative is often not very precisely known, we establish Principle 4, to cover a wide range of possibilities. It may be modified somewhat to demand good power over only certain classes of alternatives; e.g., in a test for exponentiality, one may want power only against heavy-tailed distributions.

7. <u>Regression tests</u>. These are especially suitable for situations where unknown parameters are location and scale parameters, and have been developed for tests for normality and exponentiality, particularly by Shapiro and Wilk (1965, 1972). The test is based on a regression of the order statistics of the sample against the expected values of order statistics from some canonical version of the distribution tested. Thus suppose m_i, for i from 1 to n, are the expected values of order statistics from $N(0,1)$; and let the

test be that an x-set comes from $N(\mu, \sigma^2)$, μ and σ^2 unknown. Then suppose we order $x_1 \leq x_2 \leq \cdots \leq x_n$. We have

$$E(x_i) = \mu + \sigma m_i \tag{1}$$

and a regression of x_i on m_i should be a straight line.

The x_i are correlated because they have been ordered; but using generalized least squares, an estimate of σ can be found as the estimated slope of the line. This is

$$\hat{\sigma} = \frac{\underset{\sim}{m}'\underset{\sim}{V}^{-1}\underset{\sim}{x}}{\underset{\sim}{m}'\underset{\sim}{V}^{-1}\underset{\sim}{m}} = \underset{\sim}{a}'\underset{\sim}{x} ,$$

say, where $\underset{\sim}{m}$ is the vector of m_i , $\underset{\sim}{x}$ the vector of x_i , and $\underset{\sim}{V}$ the covariance matrix of normal order statistics. The test statistic W depends essentially on a comparison of $\hat{\sigma}^2$ with s^2 , the usual sample variance. From the theoretical point of view, many problems are posed by W , both with distribution theory, even asymptotic, and by the fact that V is not known for high n ; nevertheless, in practice, Shapiro and Wilk gave approximate values of $\underset{\sim}{a}$ above, for $n \leq 50$. For $n > 50$, Shapiro and Francia (1972) replace $\underset{\sim}{m}'\underset{\sim}{V}^{-1}$ by $\underset{\sim}{m}'$; for both statistics, null distribution points are given by Monte Carlo methods. A disadvantage in the test for normality is that it needs a different $\underset{\sim}{a}$ for each n . De Wet and Venter (1973) have investigated the statistics arising when we regress $\underset{\sim}{x}$ on $\underset{\sim}{h}$ where h_i , the i-th component of $\underset{\sim}{h}$, is the inverse of the standard normal distribution evaluated at $i/(n+1)$; this is "close to" m_i but not equal to it. For theoretical purposes $\underset{\sim}{h}$ is easier to work with, and for example, de Wet and Venter have found the asymptotic distribution of the statistic corresponding to W. LaBrecque (1972) has extended the model (1) to

$$E(x_i) = \mu + \sigma m_i + \beta_2 w_2(m_i) + \beta_3 w_3(m_i) \ \ldots$$

where $w_2(m_i)$ is a quadratic and $w_3(m_i)$ a cubic in m_i, and used regression estimates of β_2 and β_3 to gives tests. Power is increased again for most alternatives, but the price is extensive computation; also one begins to get further from straightforward interpretation of the test statistics. Stephens (1975) also has investigated this model.

8. Advantages and disadvantages of W. The W type of statistic certainly satisfies Principles 1 and 2; the regression picture is informative. For the normal test, it has high power (not so overwhelmingly better than EDF statistics as at first asserted, see Stephens (1974b), but nevertheless on the whole marginally better than W^2 or A^2, and these in turn are much better than D or χ^2). On the other hand, W is more difficult to compute, and another disadvantage is the lack of mathematical theory referred to above. Of course, this may be rectified in due course. Connected with this is a greater difficulty: the W technique may not be consistent. The corresponding test for the exponential distribution above is based on a test statistic

$$W_E = \frac{n(\bar{x} - x_1)^2}{(n-1)\{\sum_i x_i^2 - n\bar{x}^2\}},$$

and Shapiro and Wilk have presented this test and given Monte Carlo percentage points. Sarkadi (1975) has recently cast doubt on the consistency of the Shapiro-Wilk tests for normality and exponentiality, though he affirms the consistency of the Shapiro-Francia (1972) test. Sarkadi gives a version of

ω_E which he states is consistent, but no percentage points. It would seem that more work is still needed on the consistency of these procedures, and indeed of other statistics.

9. Before turning to tests for exponentiality in more detail, we point out other important lines of recent research.

(a) In connection with eliminating nuisance parameters before a test can be applied, O'Reilly and Quesenberry (1973) have a method, the conditional probability integral transformation, which is elegant and of considerable mathematical interest; however, quite extensive computations are needed.

(b) Watson, and later Durbin and co-workers, and Stephens, have developed $y_n(x) = \sqrt{n}\,(F_n(x) - F(x;\theta))$ as a Fourier series; W^2 is a functional of $y_n(x)$ and its properties can be written in terms of the Fourier coefficients. Similar work was done also on U^2 and A^2 by these authors. (Durbin and Knott, 1972; Durbin, Knott, and Taylor 1975; Stephens, 1974a).

Asymptotic distributions, and also asymptotic power, can be found for certain alternatives (tending, of course, to the null as $n \to \infty$). Durbin and his co-workers also propose use of the low-order Fourier coefficients (which they call <u>components</u>) as test statistics, and show these to be more powerful, in certain circumstances, than the entire statistics (e.g., in a test of $N(0,1)$ when the alternative is $N(\gamma/\sqrt{n}, 1)$, γ a positive constant). However, in Stephens (1974a) it is shown that these circumstances are restricted, so much so that it is difficult to recommend components. For example, in a test for $N(0,1)$ against $N(\gamma_1/\sqrt{n}, 1 + \gamma_2/\sqrt{n})$, a component sensitive to the value of γ_1 might be totally unaffected by γ_2. From a practical point of view, especially when parameters must be estimated, they violate Principle 1.

For example, consider the test for exponentiality, $\alpha = 0$, β unknown; one first must calculate, from x_i , the value

$$\hat{x}_i = 1 - \exp(-x_i/\overline{x}) \ ;$$

then, for $j = 1,2,\ldots,10$, the values

$$y_j = \sqrt{2/n} \sum_{i=1}^{n} \cos(\pi j \ \hat{x}_i) \ ;$$

finally, the k-th <u>component</u> is given by

$$z_k = \sum_{j=1}^{10} b_{kj} \ y_j \ ,$$

where b_{kj} are found by methods of Durbin, Knott and Taylor (1975). (It may be that the b_{kj} given in the Durbin, Knott and Taylor paper are in error (Stephens, 1976c)).

Not only are these techniques (also that of de Wet and Venter (1973)) mathematically interesting; they also give impetus to questions concerning the purpose, efficiency, and utility of goodness-of-fit tests, and how to start measuring these concepts.

(c) On very practical lines, recent work by D'Agostino and Pearson (1973) and by Bowman and Shenton (1975) has encouraged the use, in tests for normality, of the skewness statistics b_1 and the kurtosis statistics b_2 . These are well established statistics; the first authors gave new tables of significance points, and suggested how they might be combined in an omnibus test statistic; however, they assumed independence of b_1 and b_2 , and this is not justified. Another approach to using both $\sqrt{b_1}$ and b_2 is to mark them as a point in a 2-dimensional graph; if the point falls outside given contours on the graph, normality will be rejected. This is the approach of Bowman and Shenton (1975); it provides a technique for using two statistics simultaneously to assess H_0 , although one can regard the contour as a function of b_1 and b_2 which is in effect a third statistic by which H_0 is judged. The above methods appeal to Principle 1; the practical man has a "feel" for b_1 and b_2 and, possibly, for the implications if the values lead to rejecting H_0.

10. Tests for the exponential distribution. We now return to the exponential test, to the special case where α is known. This may be taken as zero; it is convenient to replace β by θ so that H_0 is that the sample is from

$$F(x;\theta) = 1 - \exp(-\theta x)\ ,\ x > 0\ , \tag{2}$$

where the scale parameter θ is unknown.

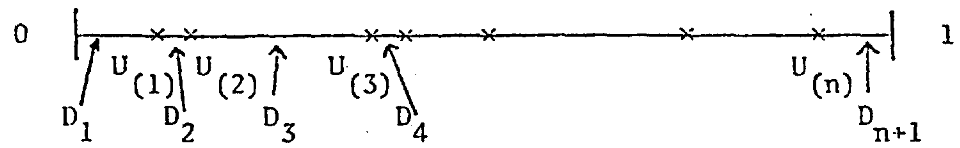
A multitude of test statistics have been posed for this important problem; not only does the distribution arise in many statistical problems, but it also appears prominently in the theory of renewal processes, in particular, the Poisson process. If events are random in time, we expect the intervals between them to be exponential. Important questions are then whether this is so, against a more general Gamma alternative, say, or whether they are exponential, but θ is changing. Some test statistics have been devised with these applications in mind.

We have already mentioned (a) EDF statistics and (b) Shapiro-Wilk W_E as available tests. Most other tests exploit some interesting connections between the exponential distribution and the uniform. These are briefly described as follows.

(a) Suppose $X_1, X_2, \ldots, X_n$ [not in order] is an exponential sample from $F(x;\theta)$; let $X_{(i)}$ be the ordered sample.

(b) Suppose $U_1, U_2, \ldots, U_n$ is a random (i.e., unordered) sample from $U(0,1)$, and let $U_{(i)}$ be the ordered sample; let D_i be the spacings between the $U_{(i)}$; i.e., $D_1 = U_{(1)}$, $D_2 = U_{(2)}-U_{(1)}$, ..., $D_n = U_{(n)}-U_{(n-1)}$, $D_{n+1} = 1-U_{(n)}$. Note that the spacings will <u>not</u> be ordered; let $D_{(i)}$ denote the ordered spacings.

(c) <u>The G transformation:</u> <u>uniforms to uniforms</u> .



The picture is obvious. Let $D_{(0)} = 0$. Let

$$D_j' = (n + 1 - j)(D_{(j)} - D_{(j-1)}) , \quad j = 1,\ldots,n+1 .$$

Then D_j' is <u>another set of unordered uniform spacings</u> (Sukhatme, 1937). One can clearly build up another <u>ordered uniform sample,</u>

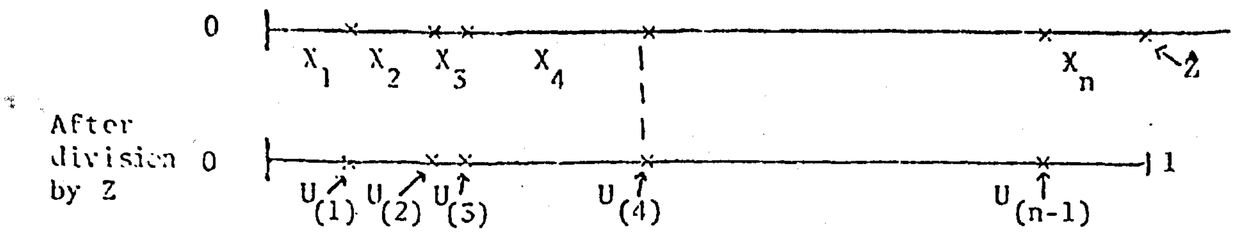
$$\underline{U'_{(j)} = \sum_{i=1}^{j} D_i'} .$$

We shall write $U' = GU$ for this transformation. Durbin (1961) showed that, loosely speaking, G makes large spacings larger and small spacings smaller. It makes a "nearly uniform" sample appear <u>further removed from uniform,</u> though it cannot be repeated indefinitely; after a certain point, an extremely non-uniform sample, i.e. very uneven spacings, is transformed to appear more uniform than before.

(d) <u>The J transformation:</u> <u>exponentials to uniforms.</u> Put the successive (unordered, exponential) X_i along a line in sequence. Let the total $\Sigma X_i = Z$. Divide by Z to get

$$U_{(j)} = \sum_{i=1}^{j} X_i/Z , \quad j = 1,\ldots,n-1 .$$

Then (well known): $U_{(j)}$ are ordered uniforms: note there are only $n - 1$ of them. We write $\underset{\sim}{U} = J\underset{\sim}{X}$.



In the case of renewal processes, the quantities $ZU_{(j)}$ shown with a cross, will be times of events (j-th blip on a screen, j-th battery failure (batteries being inserted as soon as one fails)). One can adapt the above to the situation where one observes for fixed time T_1, and obtains n observations (used batteries) in T_1, and time left over. Divide by T_1 and the $U_{(j)}$ are n uniforms.

(e) The K transformation: exponentials to uniforms.

(1) Start with X (exponentials); find $\underset{\sim}{U} = J\underset{\sim}{X}$. Note: $\underset{\sim}{U}$ are ordered uniforms; but their sizes depend on the original random ordering of the X_i.

(2) Apply G to get $\underset{\sim}{U}' = G\underset{\sim}{U}$. $\underset{\sim}{U}'$ are ordered uniforms; the combined $\underset{\sim}{U} = GJ\underset{\sim}{X}$ gives the same set of U' no matter what the original X-ordering. A formula can be found to give U_i' directly from the X_i; we write $\underset{\sim}{U}' = K\underset{\sim}{X}$ (= GJX) (Seshadri, Csorgo and Stephens, 1969).

(f) The N transformation: exponentials to exponentials. We can complete the circle by returning the U_i' to exponentials X_i', and these can be found directly from the X_i. Let $X_{(0)} = 1$;

$$X_i' = (n + 1 - i)(X_{(i)} - X_{(i-1)}) , \qquad i = 1,\ldots,n .$$

We write $\underset{\sim}{X}' = N\underset{\sim}{X}$; the X' are exponential, parameter 0, i.e. from distribution (2).

11. Exponential tests. Note that the transformations J and K reduce an exponential sample to ordered uniforms; in so doing they eliminate the unknown parameter θ. Conversely, properties of the uniform distribution can thus be used to give estimates of θ, including estimates from only the first r values of X, i.e. a censored sample. The maximum likelihood estimate from a complete sample, used in EDF tests, is $1/\bar{X}$.

Once the exponentials have been changed to uniforms, tests for uniformity can be directly applied. Sometimes this is the way the tests are proposed; at other times the tests are given in terms of the original exponentials. These connections are pointed out in order to try to see the relations between many of the proposed statistics.

12. Proposed tests for exponentiality.

The following are some proposed techniques for testing for exponentiality, roughly graded into groups.

Group A - Direct use of the EDF.

(1) EDF statistics: direct estimation of θ and use of D, V, W^2, U^2 or A^2, say. 1-tail test

(2) Srinivasan $\tilde{D}$: closely related to D above and asymptotically equivalent to it (Moore, 1973); uses Rao-Blackwell theorem for better estimate of $F(x;\theta)$. 1-tail test

(3) Finkelstein and Schafer, S^*: an EDF test, similar to W^2 in (1);

$S^* = \sum_i \delta_i$ where $\delta_i = \max[|F(X_{(i)};\hat{\theta}) - (i-1)/n|, |F(X_{(i)};\hat{\theta}) - i/n|]$.

Monte Carlo points are given for S^*. 1-tail test

Group B - Regression.

(4) Modified Shapiro-Wilk (Stephens, 1977)

$$W^* = \frac{(\Sigma x_i)^2}{n\{(n+1)\ \Sigma x_i^2 - (\Sigma x_i)^2\}}$$

distributed like W_E for $n + 1$ observations. 2-tail test

(The test for origin "known", as given by Shapiro and Wilk (1972) is strictly a composite test, for both exponential shape and a hypothesized origin; W^* is the version comparable to other tests being considered here. For the alternatives below, it is generally better than the W_E test given by Shapiro and Wilk (Stephens, 1977)).

Group C - Tests based on transformations to uniformity.

Transformations J and K below refer to those described in Section 11.

If we decide to produce uniforms by J or K, we can then follow up with appropriate tests for uniformity, e.g.

(i) EDF tests, or

(ii) tests based on spacings; much impetus to this established line of work was given by Pyke (1965);

(iii) tests based on the position of the mean of the U or U' (produced by J or K respectively) or of one order statistic $U_{(r)}$ or $U'_{(r)}$ for a chosen r;

(iv) the Shapiro-Wilk regression test. This is now based on the range over the standard deviation of the U (or U') sample.

Some tests using these techniques are:

(5) J : followed by EDF statistics in (1) above for testing for uniformity.

Lewis (1965) shows the procedure to lead to inconsistent tests; this was verified also in Seshadri, Csorgo and Stephens (1969).

(6) K : followed by EDF statistics for testing uniformity, i.e.

$$K \text{ , then } D \text{ , } V \text{ , } W^2 \text{ , } U^2 \text{ or } A^2$$ 1 tail test

These were considered by Seshadri, Csorgo and Stephens (1969), and are generally better than (6).

(7) K : followed by $\bar{U}'$, i.e. the mean of U' . Suggested by Lewis (1965): Lewis' S' is

$$S' = \bar{U}' = 2n - 2 \sum_i i \, X_{(i)}/Z$$

where $Z = \Sigma X_i$. This can be regarded as a test derived from regression of $X_{(i)}$ on i .. 2 tail test

(8) K : followed by $U'_{(r)}$ for suitable r . First suggested by Lewis (1965), and later by Tiku, Rai and Mead (1974); the latter recommended $r \simeq n/2$. 2 tail test

Group D

Other tests. Let $Z = \sum_{i=1}^{n} X_i$, $\bar{X} = Z/n$ (the X_i are the original exponentials from (2)). Let m_i be $E(X_{(i)})$ when $\theta = 1$; m_i is then $\sum_{j=1}^{i} (n - j + 1)^{-1}$.

(9) Jackson (1967):

$$T = \frac{\sum_{i=1}^{n} m_i X_{(i)}}{Z} .$$ 2-tail test

This can be regarded as derived from a regression of $X_{(i)}$ on m_i .

(10) Moran (1947, 1951):

$$M = -2 \sum_{i=1}^{n} \ln\left(\frac{X_i}{\bar{X}}\right)$$

equivalent to

$$M^* = -2 \sum_{i=1}^{n} (\ln X_i - \ln \overline{X})$$

equivalent to

$$M^{**} = -2 \sum_{i=1}^{n} \ln nD_i$$

On H_0 , M is distributed χ^2_{2n} . <u>2-tail test</u>

M is related to Bartlett's test for equal variances. Moran's test is asymptotically most powerful against the general Γ-distribution alternative (Shorack, 1972):

$$f(x) = \theta^a x^{a-1} \exp(-\theta x)/\Gamma(a) \; , \quad \underline{x > 0 \; , \; a > 0} \; .$$

It is a strong test against the Weibull alternative

$$f(x) = \theta(\theta x)^{a-1} \exp(-(\theta x)^a/a) \; , \quad \underline{x > 0 \; , \; a > 0}$$

(Bartholomew, 1957).

(11) Bartholomew (1957): $S = \sum_{i=1}^{n} (X_i/Z)^2$ equivalent to $S = \sum_{i=1}^{n} D_i^2$.

<u>1-tail test</u>

The second form of S was discussed by Moran (1947).

(12) Gail and Gastwirth (1976): this is a recently proposed test based on

$$L_n(p) = \sum_{i=1}^{r} X_{(i)}/Z$$

where $r = [np]$. They suggest $p = 0.5$.

(13) Transformation N (exponentials to exponentials), followed by Moran's M .

Considered by Epstein (1960); Fercho and Ringer (1972) gave power against Weibull alternatives, in the context of arrival times, and failure rates.

13. Power studies. We have examined the power of most of these statistics by Monte Carlo methods. Some results, for $n = 10$ and 20, and for 10% tests, are given in Tables 1 and 2.

Notes:

(a) Powers are given in percent; numbers in parentheses indicate the number (in 1000's) of M.C. samples used for the experiment.

(b) Authors usually recommend whether the test is one-tail or two-tail; this advice was followed, and no two-tail test was adapted to one-tail in order to improve the power for a specific alternative. Thus the test is performed as though the alternative were unknown; naturally one might adapt the test to one-tail against known alternatives. This might be particularly important where the test is for inter-arrival times or failure rates, and the alternative is a speeding-up of observations, or a lower (higher) failure rate.

(c) Mostly our results match those given by other authors, the exception being where those authors did adapt the tests as in (b) above.

14. Conclusions.

(a) The following all give good results:

(1) EDF using estimation of θ, and the A^2, W^2 or Finkelstein-Schafer statistic,

(2) Lewis, Jackson, Shapiro-Wilk,

(3) Moran,

(4) K-transform, using A^2 , W^2 .

(b) Groups (1) and (2) in (a) follow Principle 1 quite closely: the EDF picture is informative, the Lewis and Jackson statistics are dependent on the slopes of a plot of $X_{(i)}$ against i and m_i respectively (ignoring covariances); the Shapiro-Wilk test uses a regression on m_i , including covariance. Moran's statistic is perhaps less pictorial, but if the X_i are put sequentially along a line (not ordered), the spacings should be like uniform spacings and Moran tests this. It has the merit of known best power against some important alternatives. The K transform is less direct: though if the mean of the ordered uniforms so generated is then used as test statistic, it leads directly to Lewis' statistic.

(c) If our results for Moran are compared with those of Fercho and Ringer (1972) for the N transformation followed by Moran's statistic (there called Epstein), it seems that the direct use of Moran's statistic is preferable for the Weibull alternative considered (i.e., it is not helpful to use the N transformation).

(d) The Kolmogorov-Smirnov statistic is not good either with EDF directly or after the transformation K ; similarly for the closely connected Srinivan's $\tilde{D}$.

(e) It is still necessary to pursue research on consistency. Two main problems are involved: (1) are given tests, say for uniformity when a transformation like K has been made, themselves consistent against departures (or at least a wide class) from uniformity? (2) do several parent populations all give uniformity under one of these transformations?

(f) The test considered here has been the test that a sample is from an exponential parent population. When the problem concerns arrival times or other problems where the exponentials arrive sequentially, there will be special alternatives which will promote one test over another. It would appear that further work is needed on the efficiency of tests for this particular problem.

Acknowledgement

The author thanks Mr. H. Ghezzo for programming, Mr. Philip Giles for programming and other assistance in connection with this work, and McMaster University for a grant to support the research assistance. Gratitude is also expressed to the National Research Council of Canada, and to the U.S. Office of Naval Research, for partial support of the research.

REFERENCES

Bartholomew, D.J. (1957). Testing for departure from the exponential distribution. Biometrika, 44, 253-257.

Bowman, K.O. and Shenton, L.R. (1975). Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and b_2. Biometrika, 62, 243-250.

D'Agostino, Ralph and Pearson, E.S. (1973). Tests for departure from normality: Empirical results for the distributions of b_2 and $\sqrt{b_1}$. Biometrika, 60, 613-622.

de Wet, T. and Venter, J.H. (1973). Asymptotic distributions for quadratic forms with applications to tests of fit. Ann. Statist., 1, 380-387.

Durbin, J. (1973). Distribution theory for tests based on the sample distribution function. Philadephia; SIAM.

Durbin, J. (1975). Kolmogorov-Smirnov tests when parameters are estimated with applications to tests of exponentiality and tests on spacings. Biometrika 62, 5-22.

Durbin, J. and Knott, M. (1972). Components of the Cramér-von Mises statistics, I. J. R. Statist. Soc. B, 34, 290-307.

Durbin, J., Knott, M. and Taylor, C. C. (1975). Components of Cramér-von Mises statistics, II. J. R. Statist. Soc. B, 37, 216-237.

Epstein, B. (1960). Testing for the validity of the assumption that the underlying distribution of life is exponential. Technometrics, 2, 83-101.

Fercho, W.W. and Ringer, L.J. (1972). Small sample power of some tests of the constant failure rate. Technometrics, 14, 713-724.

Finkelstein, J.M. and Schafer, R.E. (1971). Improved goodness-of-fit tests. Biometrika, 58, 641-645.

Jackson, O.A.Y. (1967). An analysis of departures from the exponential distribution. J. R. Statist. Soc. B, 29, 540-549.

La Brecque, J.F. (1973). New goodness-of-fit procedures for the case of normality. Ph.D. thesis, SUNY at Buffalo, Department of Statistics.

Lewis, P.A.W. (1965). Some results on tests for Poisson processes. Biometrika, 52, 67-77.

Moore, D.S. (1973). A note on Srinivasan's goodness-of-fit test. Biometrika, 60, 209-211.

Moran, P.A.P. (1947). The random division of an interval - Part I. J.R. Statist. Soc. B, 9, 92-98.

Moran, P.A.P. (1951). The random division of an interval - Part II. J. R. Statist. Soc. B, 13, 147-150.

O'Reilly, F.J. and Quesenberry, C.P. (1973). The conditional probability integral transformation and applications to obtain composite chi-square goodness-of-fit tests. Ann. Statist., 1,74-83.

Pearson, E.S. and Hartley, H.O. (1972). Biometrika Tables for Statisticians, Vol. 2. Cambridge University Press.

Pettitt, A.N. and Stphens, M.A. (1976). EDF tests for the gamma distribution with application to testing for equal variances. Submitted to J. Amer. Statist. Assoc.

Pyke, R. (1965). Spacings. J. R. Statist. Soc. B, 27, 395-449.

Sarkadi, K. (1975). The consistency of the Shapiro-Francia test. Biometrika, 62, 4:5-450.

Schafer, R.E., Finkelstein, J.M. and Collins, J. (1972). On a goodness-of fit test from the exponential distribution with mean unknown. Biometrika, 59, 222-224.

Seshadri, V., Csorgo, M. and Stephens, M.A. (1969). Tests for the exponential distribution using Kolmogorov-type statistics. J. R. Statist. Soc. B, **31**, 499-509.

Shapiro, S.S. and Francia, R.S. (1972). Approximate analysis of variance test for normality. J. Am. Statist. Assoc., **67**, 215-225.

Shapiro, S.S. and Wilk, M.B. (1965). An analysis of variance test for normality. Biometrika, **42**, 591-611.

Shapiro, S.S. and Wilk, M.B. (1972). An analysis of variance test for the exponential distribution. Technometrics, **14**, 355-370.

Shorack, G.R. (1972). The best of exponentiality against gamme alternatives. J. Amer. Statist. Assoc., **67**, 213-214.

Srinivasan, R. (1970). An approach to testing the goodness of fit of incompletely specified distributions. Biometrika, **57**, 605-611.

Stephens, M.A. (1974a). Components of goodness-of-fit statistics. Annals. de L'Institut Henri Poincaré, Ser. B, **10**, 37-54.

Stephens, M.A. (1974b). EDF statistics for goodness of fit and some comparisons. J. Amer. Statist. Assoc., **69**, 730-737.

Stephens, M.A. (1975). Asymptotic properties of covariance matrics of order statistics. Biometrika, **62**, 23-28.

Stephens, M.A. (1976a). Asymptotic results for goodness-of-fit statistics with unknown parameters. Ann. Statist., 4, 357-369.

Stephens, M.A. (1976b). Goodness-of-fit for the extreme value distribution. Stanford University Technical Report, Department of Statistics, Stanford University. To appear in Biometrika.

Stephens, M.A., (1976c). Asymptotic power of EDF statistics for exponentiality against Gamma and Weibull alternatives. Technical Report, Department of Statistics, Stanford University.

Stephens, M.A. (1977). On the W test for exponentiality with origin known. Technical Report, Department of Statistics, Stanford University

Sukhatme, P.V. (1937). Tests of significance for samples of the χ^2 population with two degrees of freedom. <u>Ann. Eugen., Lond.</u>, 8, 52-56.

Sukhatme, S. (1972). Fredholm determinant of a positive definite kernel of a special type and its application. <u>Ann. Math. Statist.</u>, 43,

Tiku, M.L., Rai, K. and Mead, E. (1974). A new statistic for testing exponentiality. <u>Communications in Statistics</u>, 3, 485-493.

The distributions referred to in the table include:

Beta (a, b)	:	$f(x) = \text{const. } x^{a-1}(1-x)^{b-1}$;	$0 \leq x \leq 1$
Uniform	:	$f(x) = 1$;	$0 \leq x \leq 1$
Lognormal θ	:	$f(x) = \text{const. } \exp\{-(\ln x)^2/2\theta^2\}$;	$x \geq 0$
Weibull θ	:	$f(x) = \text{const. } x^{\theta-1} \exp(-x^{\theta})$;	$x \geq 0$
Half-normal	:	y is $N(0,1)$, x is $\|y\|$;	$x \geq 0$
Half-Cauchy	:	y is Cauchy, median zero; x is $\|y\|$;	$x \geq 0$

Test statistics are abbreviated as follows:

S - Srinivasan
F-S - Finkelstein-Schafer
S-W-S - Shapiro-Wilk, modified by Stephens
S-C-S - Seshadri-Csorgo-Stephens
L - Lewis
T-R-M - Tiku-Rai-Mead
J - Jackson
M - Moran

Table 1a. POWER OF TEST FOR EXPONENTIALITY, ORIGIN KNOWN.

Test level is 10%. The table gives the percentage of 1000M samples of size n, declared significant by the various statistics: M is given in parentheses.

n = 10	Case 4: D	W^2	V	U^2	A^2	D^+	D^-	S	F-S	S-W-S
Chi-square D.F. = 1	32(2)	35(2)	29(2)	30(2)	54(7)	42(2)	2(2)	41(4)	38(4)	33(2)
Exponential	-	-	-	-	9.5(4)	-	-	9.4(4)	10.0(4)	10.0(2)
Chi-square D.F. = 4	35(2)	40(2)	32(2)	36(2)	32(6)	1(2)	46(2)	25(4)	39(4)	33(2)
Chi-square D.F. = 6	62(2)	71(2)	60(2)	65(2)	65(6)	0(2)	74(2)	53(4)	71(4)	64(2)
Uniform (0,1)	41(3)	49(3)	48(2)	48(2)	44(7)	1(2)	55(2)	34(4)	54(4)	51(2)
Beta (1,4)	13(1)	13(1)	13(1)	13(1)	11(5)	2(1)	19(1)	9(4)	13(4)	10(2)
Log normal 1.0	18(4)	18(4)	17(2)	17(2)	16(7)	12(2)	15(2)	18(4)	18(4)	20(2)
Log normal 2.0	67(1)	70(1)	60(1)	62(1)	76(2)	74(1)	1(1)	75(1)	71(1)	72(1)
Log normal 2.4	82(1)	84(1)	77(1)	78(1)	90(2)	87(1)	1(1)	86(1)	85(1)	83(1)
Weibull 0.5	68(1)	72(1)	60(1)	62(1)	83(5)	77(1)	1(1)	70(4)	69(4)	66(2)
Weibull 2.0	66(1)	74(1)	64(1)	68(1)	71(4)	1(1)	78(1)	59(4)	78(4)	75(2)
Half normal	20(4)	22(4)	19(2)	21(2)	18(7)	1(2)	25(2)	15(4)	23(4)	21(2)
Half Cauchy	44(3)	47(3)	39(2)	41(2)	48(7)	49(2)	4(2)	49(4)	45(4)	52(2)

Table 1a. (Continued)

n = 10	Transformation K Followed By D	W^2	A^2	L	T-R-M	J	M
Chi-square D.F. = 1	38(5)	41(1)	52(5)	43(4)	42(1)	35(4)	58(4)
Exponential	9.8(4)	-	9.1(4)	9.7(4)	8.9(1)	9.3(4)	9.4(4)
Chi-square D.F. = 4	28(4)	-	54(5)	34(4)	22(1)	29(4)	38(4)
Chi-square D.F. = 6	59(4)	-	31(4)	67(4)	41(1)	57(4)	74(4)
Uniform (0,1)	42(5)	47(1)	61(4)	49(4)	42(1)	54(4)	37(4)
Beta (1,4)	10(4)	-	11(4)	11(4)	12(1)	10(4)	12(4)
Log normal 1.0	17(5)	13(1)	15(5)	17(4)	10(1)	17(4)	12(4)
Log normal 2.0	74(1)	-	77(1)	75(4)	72(1)	71(1)	73(1)
Log normal 2.4	86(1)	-	91(1)	90(1)	86(1)	86(1)	89(1)
Weibull 0.5	69(4)	-	80(4)	74(4)	70(1)	67(4)	82(4)
Weibull 2.0	65(4)	-	71(4)	77(4)	54(1)	70(4)	77(4)
Half normal	17(5)	15(1)	19(5)	20(4)	16(1)	19(4)	19(4)
Half Cauchy	51(6)	52(2)	50(5)	50(4)	43(1)	51(4)	40(4)

Table 1b. POWER OF TEST FOR EXPONENTIALITY, ORIGIN KNOWN.

Test level is 10%. The table gives the percentage of 1000M samples of size n, declared significant by the various statistics: M is given in parentheses.

n = 10	Case 4									
	D	W^2	V	U^2	A^2	D^+	D^-	S	F-S	S-W-S
Chi-square D.F. = 1	57(8)	62(8)	49(2)	51(2)	75(12)	69(2)	1(2)	61(4)	62(4)	47(2)
Exponential	-	-	-	-	9.9(4)	-	-	9.7(4)	9.4(4)	9.2(2)
Chi-square D.F. = 4	54(2)	61(2)	50(2)	54(2)	61(6)	1(2)	70(2)	49(4)	64(4)	54(2)
Chi-square D.F. = 6	91(2)	95(2)	88(2)	91(2)	95(6)	3(2)	96(2)	86(4)	95(4)	90(2)
Uniform (0,1)	68(8)	80(8)	79(2)	75(2)	79(12)	14(2)	83(2)	63(4)	83(4)	87(2)
Beta (1,4)	15(1)	16(1)	14(1)	14(1)	15(4)	3(1)	23(1)	14(4)	19(4)	18(2)
Log normal 1.0	21(3)	23(4)	22(2)	25(2)	23(7)	18(2)	18(2)	22(4)	24(4)	25(2)
Log normal 2.0	93(1)	94(1)	88(1)	89(1)	96(2)	96(1)	0(1)	93(1)	94(1)	92(1)
Log normal 2.4	99(1)	99(1)	96(1)	97(1)	99(1)	99(1)	0(1)	99(1)	99(1)	98(1)
Weibull 0.5	91(1)	93(1)	86(1)	88(1)	97(1)	95(1)	0(1)	88(2)	93(2)	85(1)
Weibull 2.0	93(1)	98(1)	93(1)	95(1)	97(1)	8(1)	97(1)	92(2)	98(2)	98(1)
Half normal	28(8)	33(8)	26(2)	28(2)	28(10)	1(2)	45(2)	23(2)	33(2)	30(1)
Half Cauchy	66(8)	69(8)	59(2)	61(2)	69(10)	73(2)	2(2)	70(2)	69(2)	73(1)

(Continued)

n = 10	Transformation K Followed By D	W^2	A^2	L	T-R-M	J	M
Chi-square D.F. = 1	60(5)	67(1)	75(5)	65(4)	58(2)	52(4)	80(4)
Exponential	9.6(4)	-	9.8(4)	9.4(4)	10.5(2)	10.9(4)	8.8(4)
Chi-square D.F. = 4	53(4)	-	59(4)	62(4)	38(2)	56(4)	71(4)
Chi-square D.F. = 6	89(4)	-	94(4)	95(4)	74(2)	91(4)	97(4)
Uniform (0,1)	75(5)	80(1)	89(5)	83(4)	73(2)	93(4)	56(4)
Beta (1,4)	16(4)	-	18(4)	18(4)	16(2)	23(4)	15(4)
Log normal 1.0	23(5)	21(1)	21(5)	18(4)	11(2)	23(4)	14(4)
Log normal 2.0	93(1)	-	96(1)	95(1)	91(1)	92(1)	94(1)
Log normal 2.4	99(1)	-	100(1)	99(1)	98(1)	98(1)	99(1)
Weibull 0.5	93(2)	-	97(2)	95(2)	86(1)	89(2)	97(2)
Weibull 2.0	92(2)	-	97(2)	98(1)	91(1)	97(2)	97(2)
Half normal	28(4)	31(2)	31(3)	34(2)	28(1)	39(2)	27(2)
Half Cauchy	75(5)	75(2)	73(3)	72(2)	64(1)	74(2)	58(2)

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 262	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Goodness of Fit With Special Reference to Tests For Exponentiality		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL REPORT
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Michael A. Stephens		8. CONTRACT OR GRANT NUMBER(s) N00014-76-C-0475
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, CA 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-267
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics and Probability Program Code 436 Arlington, VA 22217		12. REPORT DATE 9/18/78
		13. NUMBER OF PAGES 28
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release; Distribution is Unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

This report partially supported under U.S. Army Research Office Grant DAAG29-77-G-0031 and issued as Technical Report No. 22.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Goodness-of-fit; exponential distribution

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

PLEASE SEE REVERSE SIDE

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

A general review is given of recent advances in goodness-of-fit, and some principles are stated which it is desirable that a goodness-of-fit test should follow. The second part of the paper deals with a list of proposed tests for the exponential distribution, and some of the ways in which they are interconnected are demonstrated. The power of the tests is examined by Monte Carlo studies and two tables are given.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)